



MOVING OBJECT SELECTION BASED ON AN ACTIVE CURVE APPROACH

Marwen Nouri, Emmanuel Marilly, Nicole Vincent

► To cite this version:

Marwen Nouri, Emmanuel Marilly, Nicole Vincent. MOVING OBJECT SELECTION BASED ON AN ACTIVE CURVE APPROACH. IEEE International Conference on Image Processing, 2011, pp.1. hal-00637625

HAL Id: hal-00637625

<https://hal.science/hal-00637625>

Submitted on 2 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MOVING OBJECT SELECTION BASED ON AN ACTIVE CURVE APPROACH

Marwen Nouri ^{1,2} Emmanuel Marilly ², Nicole Vincent ¹

¹ LIPADE Lab, Paris Descartes University

² Alcatel-Lucent Bell Labs France

ABSTRACT

Recently, many innovating image segmentation approaches came to deal with the extraction of meaningful objects. The majority of these approaches need some user hints, such as scribbles designating the object to extract. This help to avoid a semantic gap. In the case of video object segmentation, drawing scribbles on each frame is a very long process. Scribbles drawn on one frame must be propagated automatically in the remainder of the video to make object extraction possible.

In this paper, a new framework for interactive video object selection is described. This framework is built upon active curve based scribbles propagation. Scribbles drawn by the user on the first frame are considered as curves. These curves are propagated automatically by our method on the next frames to designate the same object as the one designated by the user. A spatiotemporal optimization is introduced to ensure a coherent propagation.

Index Terms— Tracking, Interactive Methods, Video Objet, Scribbles propagation

1. INTRODUCTION

Video authoring is widely used in the context of movie production. This is a fastidious task generally done by professionals. With the raising of video sharing platforms e.g. YouTube, anyone is able to create his videos. Do-it-Yourself tools help to lower the barrier of video creation. Video creator profile is changing from professional to basic user, who needs simple tools to assist him. One of the most complicated tasks in video creation and authoring is object extraction and composition. Many interactive object extraction approaches [11, 3, 2] allow user to extract an object from an image or from a video. Some of these approaches need just few scribbles drawn by user, designating the object to be extracted. To extract a video object, one interaction possibility is to draw scribbles frame by frame. By this way the needed interaction is very intensive and therefore too boring and time consuming. Tracking systems can be helpful to reduce the user efforts.

Motion analysis and tracking are essential tasks for most video processing algorithms. Object tracking has been a popular topic in research literature since several years. Some

methods have been adapted in different approaches. We can mention optical flow for feature based approaches [9, 5, 11] and active contours for contour based approaches [7, 8]. To get an efficient system, some tracking approaches use models or manual interaction. This allows to deal with semantic gap and to remove confusions. The tracking method we are proposing takes place among methods requiring user hints. These methods are known as interactive methods. Interaction is classified into three categories, based on the tool used to select or to designate the desired object: spatiotemporal volume based, contour based and scribbles based. Spatiotemporal volume or 2D+t interaction methods allow user to add hints directly in video volume [11]. These approaches are not natural to be used and visualizing 2D+t is very confusing. In the contour based methods [3], user must define the contour of the object of interest by placing points on the object boundary. These methods allow getting high level information: a silhouette. But they require much time during the interaction step and are not easy to use for basic user. The third category of interaction methods [2] are based on scribbles drawing. Scribbles allow user to point out an object in a non-rigorous and natural way. The object is roughly designated, one or two scribbles can be sufficient for doing that. Refinement can be done interactively to get the desired selection. Based on these few scribbles, some segmentation or matting methods [3] are able to separate efficiently an object from the background.

In this paper, we proposed an interactive method, which aims at tracking a moving object. The user is asked to designate an object of interest by drawing scribbles (or curves). Then, the problem is to propagate these scribbles in next frames while designating the same object. Section 2 describes the curve propagation problem and the proposed active curve modeling. Section 3 presents and discusses the different forces we defined to govern the propagation. Experiments and results are presented in section 4.

2. SCRIBBLES PROPAGATION

Our method aims at propagating scribbles drawn on a video frame to the remainder of the sequence. For this purpose we chose an active curve modeling. First, we will explain the characteristics and the problems of propagating a curve. Then, we will expose the active curve method, its properties

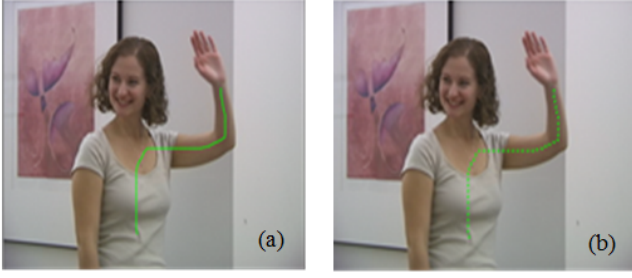


Fig. 1. (a) Example of a scribble drawn by the user. (b) A set of curves points representing discretized scribble.

and, based on it, how our problem is solved.

2.1. Curve characteristics and propagation

Propagating a hand drawn curve along the different video frames consists in tracking a set of points. Standard point based motion estimation methods, such as [4], do not allow a correct propagation of the curves points through a large number of video frames. To be correctly tracked, points must be illegible to criteria. This is the case of point of interest presented in [6]. Moreover, due to the initial curves properties (the way and where the drawing is done), tracking errors will be propagated and accumulated progressively frame after frame. This will decrease the tracking or the propagation quality. Point based tracking is sensitive to texture similarity or aperture problems. Indeed, the curves points are not geometrically independent. Processing points independently can quickly become incoherent. We suggest processing all the points as a whole, as a curve.

What are the characteristics and the functionality of the curve we want to propagate? This question is implicit to the user who wants to select specific regions to point out an object. What are the properties enabling to propagate the curve while preserving its functionality? This curve crosses several regions that compose the selected object. The curve is usually located at the middle portion of these regions as shown in Fig.1. Users drawing is usually coarse; security leads him to stay away from the objects edges.

To point out the same object in next video frames, the curve must stay in the selected regions and must be located in the middle of these areas in order to reduce the drift.

Three constraints have been identified, allowing the curve

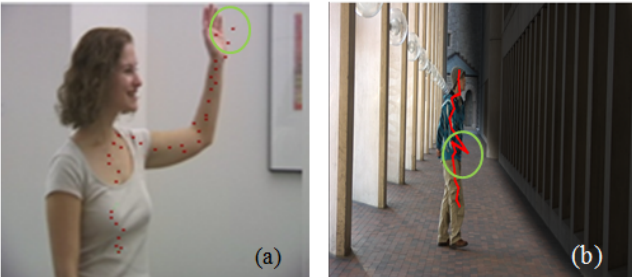


Fig. 2. (a) Example of a scribble drawn by the user. (b) A set of curves points representing discretized scribble.

propagation, that are:

- If the selected object moves, the curve has to move accordingly,
- The characteristics of crossed regions must be conserved, however, the evolution of these regions must be taken into account,
- The curve must move towards areas as locally homogeneous as possible, in order to minimize errors and to converge towards the middle of each region.

In our case, no initial assumptions can be done on the curve. As they are too global, some models like Bezier curves or splines don't seem flexible enough to enable curve propagation, considering our constraints. To perform a coherent propagation, these new constraints must be introduced to govern scribbles evolution. We define forces and the energy functional leading to position the curve on the next frames. Estimating the position of the curves in next frames is then done by an energy minimizing process based on greedy algorithm [12]. It is an active curve based approach.

3. CONSTRAINTS MODELING

To ensure a coherent propagation of scribbles points and to make them evolve while designating the same object - a set of heterogeneous areas- through video frames, we define a set of energies. First, we model the forces, associated with the curve itself and then with image data, according to the defined constraints.

3.1. Internal forces

Conventionally, the internal used forces are related to the curve bending. They are: uniformity and curvature forces. The uniformity force associated with its energy functional $E_{uniformity}$ tries to maintain the cohesion of curves points and therefore standardize the distances between each pair of successive points. The curvature energy E_{curve} influences the rigidity of the curve. These energies functional drive curve shape and define how it can change over the evolution. Weights ω_1 and ω_2 are used to let the curve evolution have a membrane like behavior or thin-plate like behavior. On each point p , the internal energy can be written as:

$$E_{internal}(p) = \omega_1 E_{uniformity}(p) + \omega_2 E_{curve}(p) \quad (1)$$

3.2. External forces

The definition of external forces is more sensitive. It is directly related to the nature of the problem. To get a coherent tracking system, which selects the same object in the successive frames, we decided to model the constraints defined previously by three forces:

- One is related to the estimated motion of each point,
- Another is related to the local color similarity,
- The last one indicates a privileged direction for propagation.

The movement of an object can be the composition of different movements from different parts of it. So, locally based **motion estimation** at each point is necessary to ensure that the overall curve is committed to the movement. Let p'_t be the optical flow [11] estimated image of a curves point p_t from frame t . We minimize the Euclidean distance between p'_t and p_{t+1} , the new points curve position in frame $t+1$. This energy formula is:

$$E_{flow}(p) = |p, p'| \quad (2)$$

Between frames at time t and $t+1$, the point must remain in the same part of the object. The area may be color uniform, but it also can be characterized by a texture. Characterization of the texture would be too time-consuming, so we chose to limit the study to the average color in a neighborhood of each curves point. We chose CIE Lab color space, which is composed from three components L, a and b. This color space allows us to distinguish between luminance and chromatic components of color. One curves point p_t will then evolve to a new position with a similar color while including some tolerance to illumination change. The color similarity energy is calculated using a weighted Euclidean distance. The average color is denoted cm . Then the energy formula is:

$$E_{color}(p) = (cm_{pt}(a) - cm_p(a))^2 + (cm_{pt}(b) - cm_p(b))^2 + 1/4 * (cm_{pt}(l) - cm_p(l))^2 \quad (3)$$

The last energy, we want to introduce, comes from the assumption that the user lacks precision when drawing the scribbles. Indeed, the user tends to draw the scribbles while trying to stay well within the middle zones of the different objects regions (which prevents him from getting out of the

object). The curve must also reproduce this effect. So to answer this third constraint, we detect the locally homogeneous region around the curve, depending on the color of a segment with a direction orthogonal to the curve in each point (Fig. 3). Our goal, starting from the state of a point on the curve at time t , is to privilege the propagation in one direction rather than the others. Therefore, we can make the point to stay in the homogeneous zone to which he belonged. This energy will push points to an area of greater homogeneity, as shown in Fig. 4. To calculate this energy, we estimate the normal to the curve at each point p_t , and calculate a similarity line segment. This segment associated with p_{curve} is the maximum segment of uniform color. That is to say, all its points colors are similar to the p_{curve} color. Maximum concerns the length of the segment. The extremities of this segment are the limit tolerance points. p_{lt} denotes the extremity which is farthest from p_t .

$$E_{privDir}(p) = |p, p'_{lt}| \quad (4)$$

The external energy is composed from three weighted terms, finally:

$$E_{global}(C) = \sum_{p \in C} (\omega_1 E_{unif}(p) + \omega_2 E_{curv}(p) + \omega_3 E_{flow}(p) + \omega_4 E_{color}(p) + \omega_5 E_{privDir}(p)) \quad (5)$$

The relative weights have been determined experimentally as usually done in active energy minimizing based systems and result on different videos are shown in the next section using the same set of weights.

4. EXPERIMENTAL RESULTS

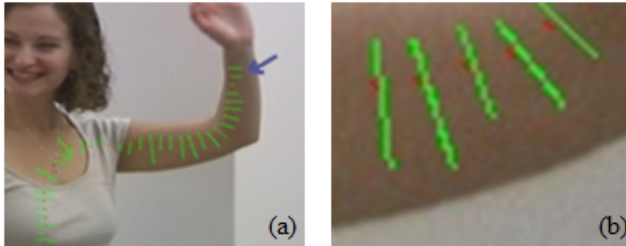


Fig. 3. (a) Similarity line segment drawn on each point of the curve. (b) Zoomed view of (a).

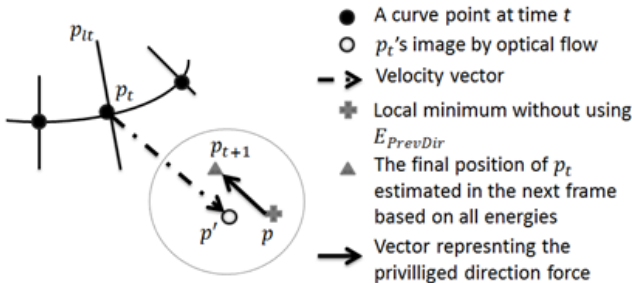


Fig. 4. Privileged direction force.

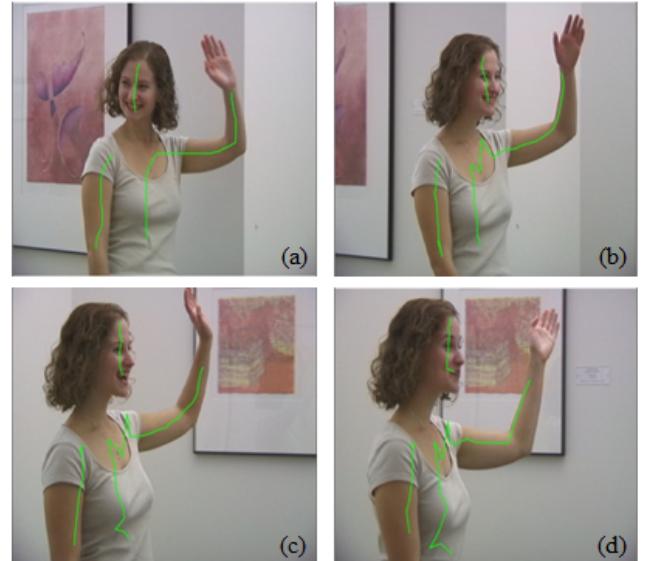


Fig. 5. (a) Three scribbles drawn at frame 1 by the user to designate the women in the video. (b)(c)(d) The respective results of propagating scribbles from frame a to frames 8, 23 and 30.

Table 1. The number of frames in which the initial selected object is still designated.

	Amira	Adam lib (29)	Walking man
OFBP	11	26	5
Our method	30	29	25

Our system proposes a graphical user interface allowing user to draw one or many scribbles at any frame of a video to point out an object. Then scribbles are propagated automatically in a real time processing. Therefore user can interact at any time to point out a new region which was not visible at the beginning of the video. Our final results are best seen in video form, though we show several still images for example in this section.

We tested our approach on three standard video sequences (640x480x30) from the literature [12, 3] (Fig. 5, 6). Our algorithm is designed to handle dynamic backgrounds as in Fig. 5 or Adam Lib sequence. Measuring objectively a success of systems like our presented one is difficult. One possible approach is based on the number of frames in which the initial designated object continues to be pointed out by our algorithm. This information is got from the final user. To compare our method, we implement an optical flow based scribbles propagator, denoted OFBP. Fig. 5a, 6a show two examples of scribbles that we want propagate. OFBP fails even on basic cases as shown on Fig. 2. Our results are shown on Fig. 5, 6 and compared to OFBP implementation in Table I. Handling optical flow errors combined with sudden illumination must be improved on further work. This can cause confusion in our system as happens after frame 25 of walking man sequence.

5. CONCLUSION

In this paper, we showed a new approach for hand drawn curve propagation. It consists in using active curves model to formulate the problem of video object selection. This allows user driven designation of objects in videos. Our contributions consist on the formulation of scribbles propagation as an active curve model and the definition of the different related forces in order to obtain a more accurate video tracking.

Our algorithm can be further improved by adding features such as texture or by using a more recent and accurate optical flow estimator. We are currently focusing on this improvement and studying the potential of our approach in two elds: video matting and human actions classification.

6. REFERENCES

[1] Anat Levin, Dani Lischinski, and Yair Weiss. A closed-form solution to natural image matting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):228242, 2008.
[2] X. Bai, J. Wang, D. Simons, and G. Sapiro. Video snapshot: Robust video object cutout using localized classifiers. In *SIGGRAPH 2009*, New York, NY, USA, 2009. ACM.

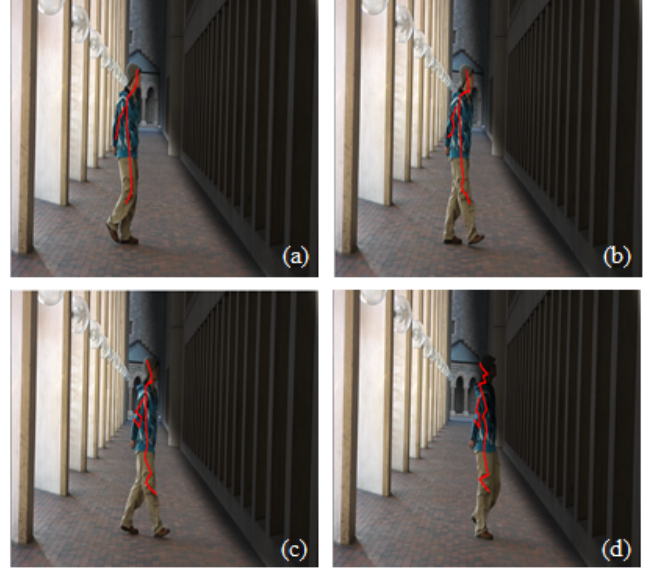


Fig. 6. (a) The user points out the man by drawing two scribbles on the first video frame of walking man sequence. (b)(c)(d) The respective results of propagating the scribbles from frame (a) to frames 7, 14 and 25.

[3] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *IJCV*, 56(3):221255, 2004.
[4] M. Irani. Multiframe correspondence estimation using subspace constraints. *IJCV*, 48(3):173194, 2002.
[5] Shi J, Tomasi C. Good features to track. *Proceedings of the IEEE CVPR*, June 1994. p. 593600.
[6] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *IJCV*, 1(4):321331, 1987.
[7] S. Lefevre and N. Vincent. Real time multiple object tracking based on active contours. In *International Conference on Image Analysis and Recognition*, volume 3212 of *Lecture Notes in Computer Sciences*, pages 606613. Springer, 2004.
[8] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of ICCV*, pages 11501157, 1999.
[9] P. Etyngier, F. Segonne, and R. Keriven. Active contour based image segmentation using machine learning techniques. in *MICCAI 2007*, ser. *Lecture Notes in Computer Science*, N. Ayache, S. Ourselin, and A. Maeder, Eds., vol. 4791. Springer, 2007, pp. 891899.
[10] P. Sand and S. Teller. Particle Video: Long Range Motion Estimation Using Point Trajectories. *IJCV*, vol. 80, May 2008, pp. 72-91.
[11] J. Wang, et al. Interactive video cutout. *SIGGRAPH05*, 24(3):585594, 2005.
[12] D.J. Williams and M. Shah. A Fast Algorithm for Active Contours and Curvature Estimation, *VIGP Computer Vision Graphics Image Process Image Understanding*, vol. 55, n 1, 1992, p. 14-26.